Aarushi Publications

B. TECH NOTES SERIES

Full Sample Chapter

Features

- As per syllabus prescribed by AICTE and Technical Universities
- Questions from prominant universities included
- Frequent realtime updates

Powered by :
AP Experts
A team of scholars & retired faculty from IIT, Roorkee

Catalogue at:

www.aarushipublications.in

BTECH NOTES SERIES

Mathematics - III (Advanced Engineering Mathematics

(As Per AICTE/Technical Universities Syllabus)

BY

AP EXPERTS

Aarushi Publications

India

Website

https://www.aarushipublications.in

Feedback

aarushipublications.kindle@gmail.com

STATISTICAL TECHNIQUES - II

TESTS OF SIGNIFICANCE

The Test of significance enable us to decide, on the basis of the results of the sample, whether

- the deviation between the observed sample statistic and the hypothetical parameter value or
- the deviation between two sample statistics is significant or might be attributed due to chance or the fluctuations of the sampling.

Null hypothesis

For applying the tests of significance, we first set up a hypothesis which is a definite statement about the population parameter called **Null hypothesis** denoted by $\mathbf{H_0}$.

Alternative hypothesis

Any hypothesis which is complementary to the null hypothesis (H_0) is called an **Alternative hypothesis denoted by $\mathbf{H_1}$.**

Example

For example, if we want to test the null hypothesis that the population has a specified mean μ_0, then we have

$$H_0: \mu = \mu_0$$

Alternative hypothesis will be

- $H_1: \mu > \mu_0$ or $\mu < \mu_0$ (two tailed alternative hypothesis).
- $H_1: \mu > \mu_0$ (right tailed alternative hypothesis *or* single tailed).
- $H_1: \mu < \mu_0$ (left tailed alternative hypothesis *or* single tailed).

Hence alternative hypothesis helps to know *whether* the test is **two tailed test** or **one tailed test**.

Critical region (region of rejection)

A region corresponding to a statistic **t**, in the sample space S which amounts to rejection of the null hypothesis H_0 is called as **critical region** or region of rejection.

Acceptance region

A region corresponding to a statistic **t**, in the sample space S which amounts to acceptance of the null hypothesis H_0 is called as **acceptance region**.

Level of significance

The probability α that a random value of the statistic t belongs to the critical region is known as the **level of significance**.

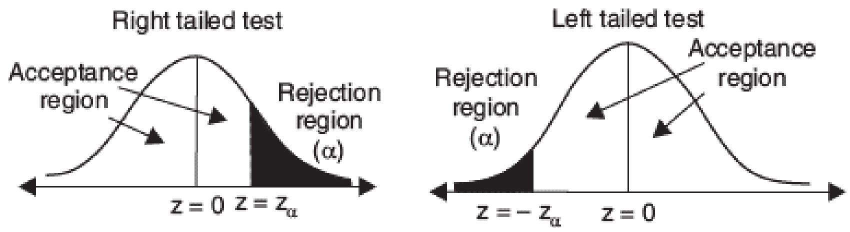
$$P(t \in \omega | H_0) = \alpha$$

Errors in sampling

- **Type I Error :** It is the error of rejecting null hypothesis when It is true. *When a null hypothesis is true*, but the difference (of mean) is significant and the hypothesis is rejected, then a Type I Error is made. The probability of making a type I error is denoted by α, *the level of significance*. In order to control the type I error, the probability of type I error is fixed at a certain level of significance α. The probability of making a correct decision is then (1 - α).
- **Type II Error:** It is the error of accepting the null hypothesis H_0 when it is false. In other words *when a null hypothesis is false*, but the difference of means is insignificant and the hypothesis is accepted, a type II error is made. The probability of making a type II error is denoted by β.

The following summary table in which H denotes the tested hypothesis may help fix the concepts of the two kinds of error:

		Truth	
		H_0 is true	H_0 is false
Decision	Accept H_0	Correct decision	Type II error (β)
Based on data	Reject H_0	Type I error (α)	Correct decision



Standard Error

The standard deviation of the sampling distribution of a statistic is known as the standard error (S.E.). It plays an important role in the theory of large samples and it forms a basis of the testing of hypothesis. If t is any statistic, for large sample

$$z = \frac{t - E(t)}{S.E(t)}$$

is normally distributed with mean 0 and variance unity.

The critical value of z at different levels of significance (α) for both single tailed and two tailed test are calculated and listed below.

Value of z_α at 5% level of significance

For two tailed test = 1.966

For right tailed = 1.645

For left tailed = -1.645

Steps in testing of statistical hypothesis

1. Null hypothesis. Set up H_0 in clear terms.
2. Alternative hypothesis. Set up H_1, so that we could decide whether we should use one tailed test or two tailed test.
3. Level of significance. Select the appropriate level of significance in advance depending on the reliability of the estimates.
4. Test statistic. Compute the test statistic $z = \frac{t - E(t)}{S.E(t)}$ under the null hypothesis.
5. Conclusion. Compare the computed value of z with the critical value z_α at level of significance (α).
   a. If $| z | > z_\alpha$, we reject H_0 and conclude that there is significant difference.
   b. If $| z | < z_\alpha$, we accept H_0 and conclude that there is no significant difference.

TESTING OF SIGNIFICANCE FOR LARGE SAMPLES

If the sample size **n > 30**, the sample is taken as **large sample.** For such sample we apply normal test, as Binomial, Poisson, chi square, etc. *are closely approximated by* normal distributions assuming the population as normal.

Under large sample test, the following are the important tests to test the significance:

- Testing of significance for single proportion.
- Testing of significance for difference of proportions.

- Testing of significance for single mean.
- Testing of significance for difference of means.
- Testing of significance for difference of standard deviations.

Testing of Significance for Single Proportion

This test is used to find the significant difference between proportion of the sample and the population. Let X be the number of successes in n independent trials with constant probability P of success for each trial.

Under H_0 test statistic

$$z = \frac{p - P}{\sqrt{PQ / n}}$$

where p = proportion of success in the sample.

Example

A coin was tossed 400 times and the head turned up 216 times. Test the hypothesis that the coin is unbiased.

Solution

H_0: The coin is unbiased i.e., P = 0.5.

H_1: The coin is not unbiased (biased); $P \neq 0.5$

Here n = 400; X = Number of success = 216

p = proportion of success in the sample = X/n = 216/400 = 0.54.

Population proportion (P) = 0.5

Q = 1 – P = 1 – 0.5 = 0.5

Under H_0, test statistic $|z| = \frac{p - P}{\sqrt{PQ / n}} = \left| \frac{0.54 - 0.5}{\sqrt{\frac{0.5x0.5}{400}}} \right| = 1.6$

As |z| = 1.6 < 1.96 [$|z| < z_\alpha$ i.e. 1.96 which is value of z at 5% significance for two tailed test.

The coin is unbiased.

Problem

A coin was tossed 400 times and the head turned up 220 times test the hypothesis that the coin

is unbiased.

Answer: The coin is unbiased

Example (JNTUK 2009, 2010, 2017, BPUT 2020 type)

In a sample of 1000 people in Karnataka 540 are rice caters and the rest are wheat caters. Can we assume that both rice and wheat are equally popular in this state at 1% level of significance?

Solution

n = 1000

p = Sample proportion of rice caters = 540/1000 = 0.54

P = Population proportion of rice caters = 1/2 = 0.5

Q = 1 - P = 1 - 0.5 = 0.5

Null Hypothesis H_0 : Both rice and wheat arc equally popular in the state.

Alternative Hypothesis H_1 : $P \neq 0.5$ (two - tailed alternative)

Test statistic $$z = \frac{p-P}{\sqrt{\frac{PQ}{n}}} = \frac{0.54-0.5}{\sqrt{\frac{0.5x0.5}{1000}}} = 2.532$$

The calculated value of z = 2.532

The tabulated value of z at 1% level of significance for two-tailed test is 2.58.

Since calculated z < tabulated z, we accept the Null Hypothesis H_0 at 1% level of significance and conclude that both rice and wheat arc equally popular in the state.

Problem (JNTUK 2009, 2010, 2013)

In a big city 325 men out of 600 men were found to be smokers. Docs this information support the conclusion that the majority of men in this city arc smokers ?

Answer: $z > z_{0.5}$ *i.e.* $2.04 > 1.645$ *(right tailed), Reject* H_0.

Testing of Significance for Difference of Proportions

Consider two samples X_1 and X_2 of sizes n_1 and n_2, respectively, taken from two different populations. Test the significance of the difference between the sample proportion p_1 and p_2. The test statistic under the null hypothesis H_0, that there is no significant difference between the two sample proportion, we have

$$z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

where $$P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}; Q = 1 - P$$

Example

A machine produced 16 defective articles in a batch of 500. After overhauling it produced 3 defectives in a batch of 100. Has the machine improved?

Solution

$p_1 = 16/500 = 0.032$; $n_1 = 500$; $p_2 = 3/100 = 0.03$; $n_2 = 100$

Null hypothesis H_0: The machine is not improved due to overhauling. Hence $p_1 = p_2$.

H_1: $p_1 > p_2$ (right tailed)

$\therefore$ $$P = \frac{0.032x500 + 0.03x100}{500 + 100} = 0.032$$

Under H_0, the test statistic $z = \dfrac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \dfrac{0.032 - 0.03}{\sqrt{(0.032)(0.968)\left(\frac{1}{500} + \frac{1}{100}\right)}} = 0.104$

We see that the calculated value of $|z| < 1.645$, the significant value of z at 5% level of significance (for right tailed test), H_0 is accepted, i.e., the machine has not improved due to overhauling.

Example (UTU 2013, 5 marks)

Before an increase in excise duty on tea, 800 people out of 1000 persons were found to be tea drinkers. After an increase in duty 800 persons were known to be tea drinkers in a sample of 1200. Do you think that there has been significant decrease in the consumption of tea after the increase in excise duty?

Solution

Given $n_1 = 1000$; $n_2 = 1200$

p_1 = proportion of tea drinkers before increase in excise duty = 800/1000 = 0.8

p_2 = proportion of tea drinkers after increase in excise duty = 800/1200 = 0.67

$$P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{1000x0.8 + 1200x0.67}{1000 + 1200} = 0.73$$

$\therefore$ Q = 1 - P = 1 0.73 = 0.27

H_o: $p_1 = p_2$ (i c., there is no significant difference in the consumption of tea before and after the increase in excise duty)

H_1 : $p_1 > p_2$ (i.e., Right tailed test)

Under test statistic

$$z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0.8 - 0.67}{\sqrt{0.73(0.27)\left(\frac{1}{1000} + \frac{1}{1200}\right)}} = 6.8$$

From table, $z_{0.05}$ = 1.645 (right tailed)

Since z > 1.645, H_o is rejected at 5% level of significance level of significance, (i.e., there is no significant difference between the people consuming tea before and after increase in excise duty)

Problem

In a random sample of 100 men taken from village A, 60 were found to be consuming alcohol. In another sample of 200 men taken from village B, 100 were found to be consuming alcohol. Do the two villages differ significantly in respect to the proportion of men who consume alcohol? The table value of z at 5% level is 1.96 (two tailed test).

Answer: z = 1.64 < 1.96, H_0 will be accepted.

Testing of Significance for Single Mean

To test whether the given sample of size n has been drawn from a population with mean μ, i.e. to test whether the difference between the sample mean and the population mean is significant. Under the null hypothesis, there is no difference between the sample mean and population mean.

The test statistic is $z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$, where σ is the standard deviation of the population.

Example

A normal population has a mean of 6.8 and standard deviation of 1.5. A sample of 400 members gave a mean of 6.75. Is the difference significant?

Solution

H_0: There is no significant difference between x and μ.

H_1: There is significant difference between x and μ.

Given μ = 6.8, σ = 1.5, $\bar{x}$ = 6.75 and n = 400

$$|z| = \left|\frac{6.75-6.8}{1.5/\sqrt{900}}\right| = 0.67$$

As the calculated value of | z | < z_α = 1.96 at 5% level of significance, H_0 is accepted, i.e., there is no significant difference between x and μ.

Example (JNTUH 2018, 10 marks)

A sample of 400 items is taken from a normal population whose mean is 4 and variance 4. If the sample mean is 4.45, can the samples be regarded as a simple sample?

Solution

Null Hypothesis (H_0) : Sample can not be regarded as having been drawn from the population with mean 4.

Alternative Hypothesis (H_1): Sample be regarded as having been drawn from population with mean 4.

Population mean (p) = 4. Sample mean ($\bar{X}$) = 4.45

variance (σ^2) = 4 ⇒ σ = 2

$$z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} = \frac{4.45-4}{2/\sqrt{400}} = 4.5$$

As the calculated value of | z | > z_α = 1.96 at 5% level of significance, H_0 is rejected i.e. sample can not be regarded as laying been drawn from the population with mean 4.

Problem

A simple sample of 1,000 members is found to have a mean 3.42 cm. Could it be reasonably regarded as a simple sample from a large population arose mean is 3.30 cm & S.D is 2.6 cm? Given $z_{0.05}$ = 1.96.

Answer: $z < z_{0.05}$ i.e. 1.46 < 1.96. H_0 is accepted. So we say that the sample is drawn from normal population.

Example (JNTUH 2005, 2019, 5 marks)

Assuming that σ = 20.0, how large a random sample be taken to assert with probability 0.95 that the sample mean will not differ from the true mean by more than 3.0 points?

Solution

$$z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

Given $z_{0.5} = 1.96$; ; $\sigma = 20$

Putting these values in above formula

We get $n = 171$

Example (JNTUH 2010, 2013, 2015, 2017, 2019, 5 marks)

A normal population has a mean of 0.1 and standard deviation of 2.1. Find the probability that mean of a sample of size 900 will be negative.

Solution

Given $\mu = 0.1$; $\sigma = 2.1$; $n = 900$

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{\bar{x} - 0.1}{2.1/\sqrt{900}} = \frac{\bar{x} - 0.1}{0.07}$$

Now $P(\bar{x} < 0) = P(0.1 + 0.07z < 0)$

$$= P\left(z < -\frac{0.1}{0.07}\right) = P\left(z < -\frac{10}{7}\right) = P\left(z < -1.43\right)$$

$$= P(z > 1.43) = 0.5 - P(0 < z < 1.43)$$

$= 0.5 - 0.4236$ [From table]

$= 0.0764$

Example (JNTUH 2005, 2019, 5 marks)

In a random sample of 60 workers, the average time taken by them to get to work is 33.8 minutes with a standard deviation of 6.1 minutes. Can we reject the null hypothesis μ =32.6 minutes in favour of alternative null hypothesis μ > 32.6 at α = 0.025 level of significance.

Solution

Null Hypothesis $H_0 = \mu = 32.6$

Alternative hypothesis $H_1 = \mu > 32.6$

Level of significance: $\alpha = 0.025$

The test statistic

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{33.8 - 32.6}{6.1 / \sqrt{60}} = 1.5238$$

Tabulated value of z at 0.025 level of significance is 2.58.

Hence, we see that $z < z_{0.025}$ i.e. $z < 2.58$.

The null hypothesis H_0 will be accepted.

Test of Significance for Difference of Means of Two Large Samples

Let $\bar{x}_1$ be the mean of a sample of size n_1 from a population with mean μ_1, and variance σ_1. Let $\bar{x}_2$ be the mean of an independent sample of size n_2 from another population with mean μ_2 and variance σ_2.

The test statistic is given by

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Example

Intelligence test of two groups of boys and girls gave the following results:

	Means	*SD*	*Size*
Girls	*75*	*8*	*60*
Boys	*73*	*10*	*100*

Is the difference in mean scores significant? Also test for SD.

Solution

Null hypothesis H_0: There is no significant difference between mean scores, i.e., $\bar{x}_1 = \bar{x}_2$.

H_1: $\bar{x}_1 \neq \bar{x}_2$

Under the null hypothesis

$$|z| = \left| \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \right| = \left| \frac{75 - 73}{\sqrt{\frac{8^2}{60} + \frac{10^2}{100}}} \right| = 1.3912$$

As the calculated value of $|z| < 1.96$, the significant value of z at 5% level of significance, H_0 is accepted i.e., there is no significant difference between mean scores.

Test of Significance for the Difference of Standard Deviations of two large samples

If σ_x and σ_2 are the standard deviations of two independent samples, then under the null hypothesis HQ: ox = o2, i.e., the sample standard deviations don't differ significantly, the statistic

When σ_1 and σ_2 are population standard deviations, then for large sample size

$$z = \frac{\sigma_1 - \sigma_2}{\sqrt{\frac{\sigma_1^2}{2n_1} + \frac{\sigma_2^2}{2n_2}}}$$

Example

Random samples drawn from two countries gave the following data relating to the heights of adult males:

	Country A	*Country B*
Mean height (inches)	*67.42*	*67.25*
Standard deviation	*2.58*	*2.50*
Number of samples	*1000*	*1200*

(i) Is the difference between the means significant?

(ii) Is the difference between the standard deviations significant?

Solution

(i) Null hypothesis: $H0 = \mu_1 = \mu_2$ i.e., sample means do not differ significantly.

Alternative hypothesis: H_1: $\mu_1 \neq \mu_2$ (two tailed test)

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{67.42 - 67.25}{\sqrt{\frac{2.58^2}{1000} + \frac{2.50^2}{1200}}} = 1.56$$

Since $|z| < 1.96$ we accept the null hypothesis at 5% level of significance.

(ii) H_0: $\sigma_1 = \sigma_2$ i.e., the sample standard deviations do not differ significantly.

Alternative hypothesis: $H_1 = \sigma_1 \neq \sigma_2$ (two tailed)

$$|z| = \left| \frac{\sigma_1 - \sigma_2}{\sqrt{\frac{\sigma_1^2}{2n_1} + \frac{\sigma_2^2}{2n_2}}} \right| = \left| \frac{2.58 - 2.50}{\sqrt{\frac{2.58^2}{2x1000} x \frac{2.50^2}{2x1200}}} \right| = 1.0387$$

Since | z | < 1.96 we accept the null hypothesis at 5% level of significance.

STUDENT T-DISTRIBUTION FOR SMALL SAMPLES (N ≤ 30)

This t-distribution is used when sample size is ≤ 30 and the population standard deviation is unknown.

t-statistic is defined as

$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}}$$

where standard deviation of sample

$$s = \sqrt{\frac{\sum (X - \bar{X})^2}{n - 1}}$$

t-test of significance of the mean of a random sample

H_0: There is no significant difference between the sample mean $\bar{x}$ and the population mean μ, i.e., we use the statistic

$$t = \frac{\bar{X} - \mu}{s / \sqrt{n}} \text{ where } \bar{X} \text{ is mean of the sample.}$$

and $$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2 \text{ with degree of freedom (n - 1).}$$

If calculated t value is such that | t | < t_α the null hypothesis is accepted. | t | > t_α, H_0 is rejected.

Example

The lifetime of electric bulbs for a random sample of 10 from a large consignment gave the following data:

Item	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
Life in "000" hrs	*4.2*	*4.6*	*3.9*	*4.1*	*5.2*	*3.8*	*3.9*	*4.3*	*4.4*	*5.6*

Can we accept the hypothesis that the average lifetime of a bulb is 4000 hrs? For df = 9, $t_{0.05}$ *= 2.26.*

Solution

H_0: There is no significant difference in the sample mean and population mean. i.e., μ = 4000 hrs.

$$\bar{X} = \frac{\sum X}{n} = \frac{44}{10} = 4.4$$

X	4.2	4.6	3.9	4.1	5.2	3.8	3.9	4.3	4.4	5.6
$X-\bar{X}$	-0.2	0.2	-0.5	-0.3	0.8	-0.6	-0.5	-0.1	0	1.2
$(X-\bar{X})^2$	0.04	0.04	0.25	0.09	0.64	0.36	0.25	0.01	0	1.44

From table, $\sum(X-\bar{X})^2 = 3.12$

$$s = \sqrt{\frac{\sum(X-\bar{X})^2}{n-1}} = \sqrt{\frac{3.12}{9}} = 0.589$$

$$t = \frac{\bar{X}-\mu}{s/\sqrt{n}} = \frac{4.4-4}{0.589/\sqrt{10}} = 2.123$$

Degree of freedom = n - 1 = 10 - 1 = 9

Given at df = 9, $t_{0.05}$ = 2.26.

Since the calculated value of t is less than table $t_{0.05}$.

∴ The hypothesis μ = 4000 hrs is accepted, i.e., the average lifetime of bulbs could be 4000 hrs.

Example (GTU NM 2020, 5 marks)

The heights of 9 males of a given locality are found to be 45, 47, 50, 52, 48, 47, 49, 53, 51 inches. Is it reasonable to believe that the average height is differ significantly from assumed mean 47.5 inches? (Given that for 5% level of significance t for 7 d. f is 2.365, for 8 d. f. is 2.306 and for 9 d. f. is 2.262)

Solution

The null hypothes is H_0: μ = 47.5

Alternative hypothesis H_1: $\mu \neq 47.5$

$$\bar{X} = \frac{45+47+50+48+47+49+53+51+52}{9} = 49.11$$

X	45	47	50	52	48	47	49	53	51	Sum
$X-\bar{X}$	-4.11	-2.11	0.89	2.89	-1.11	-2.11	-0.11	3.89	1.89	

$(X-\bar{X})^2$	16.89	4.45	0.79	8.35	1.23	4.45	1.21	15.13	3.57	56.07

$$s=\sqrt{\frac{\sum(X-\bar{X})^2}{n-1}}=\sqrt{\frac{56.07}{9}}=2.49$$

$$t=\frac{\bar{X}-\mu}{s/\sqrt{n}}=\frac{49.11-47.5}{2.49/\sqrt{10}}=1.89$$

Degree of freedom = n - 1 = 10 - 1 = 9

Given at df = 9, $t_{0.05}$ = 2.306.

Since the calculated value of t is less than table $t_{0.05}$.

∴ The null hypothes H_0 is μ = 47.5 will be accepted.

Example

A sample of 20 items has mean 42 units and standard deviation 5 units. Test the hypothesis that it is a random sample from a normal population with mean 45 units.

Solution

H_0: There is no significant difference between the sample mean and the population mean. i.e., μ = 45 units

H1: μ ≠ 45 (Two tailed test)

Given: n = 20, X = 42, S = 5; df = 20 - 1 = 19

$$s^2=\frac{n}{n-1}S^2=\left(\frac{20}{20-1}\right)(5)^2=26.31\Rightarrow s=5.129$$

$$t=\frac{\bar{X}-\mu}{s/\sqrt{n}}=\frac{42-45}{5.129/\sqrt{20}}=-2.615;|t|=2.615$$

The tabulated value of t at 5% level for 19 d.f. is $t_{0.05}$ = 2.09.

Since | t | > $t_{0.05}$, the hypothesis H_0 is rejected, i.e., there is significant difference between the sample mean and population mean. i.e., the sample could not have come from this population.

t-test for difference of means of two small samples

H_0: The samples have been drawn from the normal population with means μ_1 and μ_2, i.e., H_0: $\mu_1 \neq \mu_2$.

Let $\bar{X},\bar{Y}$ be their means of the two samples.

Under this H_0 the test of statistic t is given by

$$t = \frac{(\bar{X} - \bar{Y})}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

and df = $n_1 + n_2 - 2$

If the two sample's standard deviations s_1, s_2 are given, then we have

$$s^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}$$

Example

Two samples of sodium vapor bulbs were tested for length of life and the following results were obtained:

	Size	*Sample mean*	*Sample SD*
Type I	*8*	*1234 hrs*	*36 hrs*
Type II	*7*	*1036 hrs*	*40 hrs*

Is the difference in the means significant to generalize that Type I is superior to Type II regarding length of life? $t_{0.05}$ *at df 13 is 1.77 (one tailed test).*

Solution

$H_0: \mu_1 = \mu_2$ i.e., two types of bulbs have same lifetime.

$H_1: \mu_1 > \mu_2$ i.e., type I is superior to Type II.

$$s^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2} = \frac{8(36)^2 + 7(40)^2}{8 + 7 - 2} = 1659.076 \Rightarrow s = 40.7317$$

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{1234 - 1036}{40.1480\sqrt{\frac{1}{8} + \frac{1}{7}}} = 18.1480$$

df = 8 + 7 - 2 = 13

Given $t_{0.05}$ at df 13 is 1.77 (one tailed test).

Since calculated $|\,t\,| > t_{0.05}$, H_0 is rejected, i.e. H_1 is accepted.

$\therefore$ Type I is definitely superior to Type II.

Example (DU 2003, JNTUH 2019, 5 marks)

The mean life of a sample of 10 electric bulbs was found to be 1456 hours with S.D. of 423 hours. A second sample of 17 bulbs chosen from a different batch showed a mean life of 1280

hours with S.D. of 398 hours. Is there a significant difference between the means of two batches?

Solution

Given that

$n_1 = 10$; $\bar{x}_1 = 1456$; $s_1 = 423$

$n_2 = 17$; $\bar{x}_2 = 1280$; $s_2 = 398$

Null hypothesis H_0: $\mu_1 = \mu_2$ i.e. there is no significant difference between the means of two samples.

Alternative hypothesis H_1: $\mu_1 \neq \mu_2$ (Two tailed test)

$$S = \sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}} = \sqrt{\frac{10x423^2 + 17x398^2}{10 + 17 - 2}} = 423.42$$

Standard error of difference

$$SE(\bar{x}_1 - \bar{x}_2) = S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = 423.42\sqrt{\left(\frac{1}{10} + \frac{1}{17}\right)} = 168.52$$

Test statistics $t = \dfrac{\bar{x}_1 - \bar{x}_2}{SE(\bar{x}_1 - \bar{x}_2)} = \dfrac{1456 - 1280}{168.52} = 1.04$

Level of significance: Take $\alpha = 0.05$

The tabulated or critical value of $t_{0.05} = 2.06$ (for df = 10 + 17 - 2 = 25)

The calculated value of $|t| = 1.04 < t_{0.05}$. Hence null hypothesis H_0 will be accepted.

CHI-SQUARE (χ^2) TEST

When a coin is tossed 200 times, the theoretical considerations lead us to expect 100 heads and 100 tails. But in practice, these results are rarely achieved. The quantity χ^2 (a Greek letter, pronounced as **chi-square**) describes the magnitude of discrepancy between theory and observation.

If O_i (i = 1, 2, ..., n) is a set of observed (experimental) frequencies and E_i (i = 1, 2, ..., n) is the corresponding set of expected (theoretical or hypothetical) frequencies, then, χ^2 is defined as

$$\chi^2 = \sum_{i=1}^{n}\left[\frac{(O_i - E_i)^2}{E_i}\right]$$

Degree of freedom

The number of degrees of freedom is the total number of observations less the number of independent constraints imposed on the observations. Degrees of freedom (difference) are usually denoted by **ν** (the letter 'nu' of the Greek alphabet).

Example

If we have to choose any four numbers whose sum is 50, we can exercise our independent choice for any three numbers only, the fourth being 50 minus the total of the three numbers selected. Thus, though we were to choose any four numbers, our choice was reduced to three because of one condition imposed.

There was only one restraint on our freedom and our degrees of freedom were (n - 1) = 4 – 1 = 3. If two restrictions are imposed, our freedom to choose will be further curtailed and degrees of freedom will be 4 – 2 = 2.

Example (PTU 2011, AKTU 2019, 3.5 marks)

A die is thrown 270 times and the results of these throws are given below:

Number appeared on the die	*1*	*2*	*3*	*4*	*5*	*6*
Frequency	*40*	*32*	*29*	*59*	*57*	*59*

Test whether the die is biased or not. Tabulated value of χ^2 at 5% level of significance for *d.f.* = *5 is 11.09.*

Solution

Null hypothesis H$_0$: Die is unbiased.

Under this H$_0$, the expected frequencies for each digit is 276/6 = 46.

To find the value of χ^2

O_i	40	32	29	59	57	59
E_i	46	46	46	46	46	46
$(O_i - E_i)^2$	36	196	289	169	121	169

Now $$\chi^2 = \frac{\sum (O_i - E_i)^2}{E_i} = \frac{980}{46} = 21.30$$

Tabulated value of χ^2 at 5% level of significance for (n - 1) = 6 – 1 = 5 d.f. (degree of freedom) is 11.09. Since the calculated value of $\chi 2$ = 21.30 > 11.07 the tabulated value, H$_0$ is rejected. i.e., die is not unbiased or die is **biased**.

Example (GTU NM 2020, 5 marks)

In an experiment on immunization of cattle from tuberculosis the following result were obtained:

	Affected	*Unaffected*
Inoculated	*12*	*26*
Not inoculated	*16*	*6*

Examine the effect of vaccine in controlling the susceptibility to tuberculosis. (Given that for 5% level of significance χ^2 for 1 d. f is 3.841, for 2 d. f. is 5.99 and for 3 d. f is 7.815).

Solution

We set up the hypothesis that vaccine has no effect in controlling susceptibility to tuberculosis. On this hypothesis the **expected frequencies** are as shown in the following table:

	Affected	Unaffected	Total
Inoculated	38 x 28/60 = 18	38 - 18 = 20	38
Not inoculated	28 - 18 = 10	22 - 10 = 12	22
Total	28	32	60

$$\chi^2 = \frac{(12-18)^2}{18} + \frac{(26-20)^2}{20} + \frac{(16-10)^2}{10} + \frac{(6-12)^2}{12} = 10.44$$

Now the tabulated value of 5% level of significance for one degree of freedom is found to be 3.841. Since the calculated value is greater than this value, the hypothesis is wrong and consequently the variance is effective in controlling susceptibility to tuberculosis.

Example (UTU 2013, 2016, 10 marks):

The following frequency distribution gives the frequencies of seeds in a pea breeding experiment:

Round & yellow	*Wrinkled & yellow*	*Round & green*	*Wrinkled & green*	*Total*
315	*101*	*108*	*32*	*556*

Theory predicts that the frequencies should be 9:3:3:1. Examine the correspondence between theory and experiment.

Solution

Taking the hypothesis that the theory fits well into the experiment, the expected frequencies are respectively

$$\frac{9}{16}x556, \frac{3}{16}x556, \frac{3}{16}x556, \frac{1}{16}x556$$

i.e. 313, 104, 104, 35.

Thus

$$\chi^2 = \sum\frac{(O-E)^2}{E} = \frac{(315-313)^2}{313} + \frac{(101-104)^2}{104} + \frac{(108-104)^2}{104} + \frac{(32-35)^2}{35} = 0.513$$

O	315	101	108	32
E	313	104	104	35

Degrees of freedom = 4-1=3

Table value of χ^2 for 3 d.f. at 5% level of significance = 7.815.

Since the calculated value of χ^2 is much less than the table value, the hypothesis may be accepted. Hence there is much correspondence between theory and experiment.

Example (GTU NM 2020, 7 marks)

Five coins are tossed 3200 times and the following results are obtained:

No. of heads	0	1	2	3	4	5
Frequency	80	570	1100	900	500	50

If χ^2 for 5 d.f at 5% level of significance be 11.07, test the hypothesis that the coins are unbiased.

Solution

Let H_0 : Coins are unbiased that is, p = q = 1/2

Apply binomial probability distribution to get the expected number of heads as follows:

Expected number of heads = $n\,{}^nC_r p^r q^{n-r}$ = $3200\,{}^5C_0\left(\frac{1}{2}\right)^r\left(\frac{1}{2}\right)^{5-r} = 3200\,{}^5C_0\left(\frac{1}{2}\right)^5$

The table is given below:

O	E	$(O - E)^2$	$(O - E)^2/E$
80	100	400	4.00
570	500	4900	9.80
1100	1000	10,000	10.00
900	1000	10,000	10.00
500	500	0	0.00

50	100	2500	25.00
		Total	58.80

Since $\chi^2 = 58.80$ is more than its critical value $\chi^2 = 11.07$ for df = 6 - 1 = 5 and $\alpha = 0.05$, the null hypothesis is rejected.

Problem

The theory predicts the proportion of beans in the four groups A, B, C and D should be 9:3:3:1. In an experiment among 1600 beans, the numbers in the four groups were 882, 313, 287 and 118. Does the experimental result support the theory ? (χ^2 for 3 d.f. at 5% level = 7.815).

Answer: $\chi^2 = 4.7266 <$ table value, hypothesis may be accepted.

Problem

The following table shows the distribution of digits in numbers chosen at random from a telephone directory:

Digits	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
Frequency	*1026*	*1107*	*997*	*966*	*1075*	*933*	*1107*	*972*	*964*	*853*

Test whether the digits may be taken to occur equally frequently in the directory. The tabulated value of χ^2 at 5% level of significance for 9 difference is 16.919.

Answer: H_0 is rejected. The digits taken in the directory do not occur equally frequently.

Example (UTU 2015, 5 marks)

The weight of a drug produced by Ganga Pharmaceutical Co. follows normal distribution. The specified variance of the weight of the drug of this population is 0.25 mg. The quality engineer of the firm claims that the variance of the weight of the drug does not differ significantly from the specified variance of the weight of the drug of the population. So, the purchase officer of the Alpha Hospital who places order for that drug with the Ganga Pharmaceutical Co. has selected a random sample of 12 drugs. The variance of the weight sample is found to be 0.49 mg. Verify the intuitions of the quality manager of Ganga Pharmaceutical Co. at a significance level of 0.10 using Chi Square. (Tabulated chi square value is 19.675 with 11 dof).

Solution

Weight of the drug follows normal distribution. We also have population variance of the weight of drugs $\sigma^2 = 0.25$ mg; sample size, n = 12; sample variance of the weight of drugs, $S^2 = 0.49$ mg; and significance level, $\alpha = 0.10$.

Null and alternate hypotheses are:

H_0: $\sigma^2 = 0.25$

H_1: $\sigma^2 \neq 4\ 0.25$

It is a two tailed test.

The chi-square statistic to test the variance is:

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} = \frac{(12-1)x0.49}{0.25} = 21.56$$

df = 12 -1 = 11, $\chi^2_{0.10} = 19.675$

We see $\chi^2 > \chi^2_{0.10}$

Hence, null hypothesis H_0 will be rejected.

Problem

The weight of cement bags produced by Vignesh Cement Company follows normal distribution. The quality assistant at the final inspection section of the company feels that the variance of the weight of the cement bags has increased from a specified maximum variance of 0.64 kg which will lead to customer complaints. Hence, he has selected a sample of 8 cement bags and found that the variance of the sample is 0.36 kg. Check the intuition of the quality assistant at a significant level of 0.01. Given $\chi^2_{0.01} = 1.239$.

Answer: $\chi^2 = 3.938$, *H_0 will be accepted. Left tailed test.*

F TEST

In testing the significance of the difference of two means of two samples, we assumed that the two samples came from the same population or populations with equal variance. The object of the F-test is to discover whether two independent estimates of population variance differ significantly or whether the two samples may be regarded as drawn from the normal populations having the same variance. Hence before applying the t-test for the significance of the difference of two means, we have to test for the equality of population variance by using the F-test.

To test whether these estimates, s_1^2 and s_2^2, are significantly different or if the samples may be regarded as drawn from the same population or from two populations with same variance σ^2, we set-up the null hypothesis

H_0: $\sigma_1^2 = \sigma_2^2 = \sigma^2$,

i.e., the independent estimates of the common population do not differ significantly.

To carry out the test of significance of the difference of the variances we calculate the test statistic $F = s_1^2 / s_2^2$ the Numerator is greater than the Denominator, i.e., $s_1^2 > s_2^2$.

If the calculated value of F exceeds $F_{0.05}$ for $(n_1 - 1)$, $(n_2 - 1)$ degrees of freedom given in the table, we conclude that the ratio is significant at 5% level.

Example (BPUT 2020, 16 marks)

In a sample of 8 observations, the sum of squared deviation of items from the mean was 94.5. In other samples of 10 observations, the value was found by 101.7. Test whether the difference is significant at 5% level of significance?

Solution

H_0: $s_1 = s_2$; H_1: $s_1 \neq s_2$

Given that $n_1 = 8$; $\sum(x_1 - \bar{x}_1)^2 = 94.5$

$n_2 = 10$; $\sum(x_2 - \bar{x}_2)^2 = 101.7$

Now $$s_1^2 = \frac{\sum(x_1 - \bar{x}_1)^2}{n_1 - 1} = \frac{94.5}{7} = 13.5$$

and $$s_2^2 = \frac{\sum(x_2 - \bar{x}_2)^2}{n_2 - 1} = \frac{101.7}{9} = 11.3$$

$$F = \frac{s_1^2}{s_2^2} = \frac{13.5}{11.3} = 1.195$$

df = 7 and 9, $F_{0.05} = 3.29$

The calculated value of F is less than table value. Hence, we accept the hypothesis and conclude that the difference in the variances of two samples is not significant at 5% level.

Problem

Two independent sample of sizes 7 and 6 had the following values:

Sample A	*28*	*30*	*32*	*33*	*31*	*29*	*34*
Sample B	*29*	*30*	*30*	*24*	*27*	*28*	

Examine whether the samples have been drawn from normal populations having the same variance.

Solution

H0: The variance are equal. i.e., $\sigma_1^2 = \sigma_2^2$ i.e., the samples have been drawn from normal populations with same variance.

H1: $\sigma_1^2 \neq \sigma_2^2$

X_1	$X_1-\bar{X}_1$	$(X_1-\bar{X}_1)^2$	X_2	$X_2-\bar{X}_2$	$(X_2-\bar{X}_2)^2$
28	-3	9	29	1	1
30	-1	1	30	2	4
32	1	1	30	2	4
33	2	4	24	-4	16
31	0	0	27	-1	1
29	-2	4	28	0	0
34	3	9			
		28			26

$$\bar{X}_1 = 31; n_1 = 7; \sum(X_1-\bar{X}_1)^2 = 28$$

$$\bar{X}_2 = 28; n_2 = 6; \sum(X_2-\bar{X}_2)^2 = 26$$

$$s_1^2 = \frac{\sum(X_1-\bar{X})^2}{n_1-1} = \frac{28}{6} = 4.666$$

$$s_2^2 = \frac{\sum(X_2-\bar{X})^2}{n_2-1} = \frac{26}{5} = 5.2$$

$$F = \frac{s_1^2}{s_2^2} = \frac{4.666}{5.2} = 0.9$$

$\therefore \quad s_2^2 > s_1^2$

The tabulated value of F at df1 = 6 – 1 and df2 = 7 – 1 difference for 5% level of significance is 4.39. The calculated value of F is less than the tabulated value of F. H_0 is accepted. Hence, there is no significant difference between the variance. The samples have been drawn from the normal population with same variance.

Problem

Two independent samples of sizes 9 and 8 gave the sum of squares of deviations from their respective means as 160 and 91 respectively. Can the samples be regarded as drawn from the normal populations with equal variance? Given: F0.05 (8, 7) = 3.73 and F0.05 (7, 8) = 3.50

Answer: $F_{cal} = 3.029 < F_{0.05}(8, 7) = 3.73$ *(given)*

The Null Hypothesis~ H_0 *is accepted. The samples may be regarded as drawn from the normal population with equal variance*

ASSIGNMENT

Q.1. (JNTUH 2018, 2 marks): Write the conditions of validity of χ^2 test.

Q.2. (JNTUH 2018, 2 marks): Construct sampling distribution of means for the populations 3, 7, 11, 15 by drawing samples of size two without replacement. Determine (i) μ (ii) σ (iii) sampling distribution of means.

Q.3. (JNTUH 2018, 2 marks): Discuss types of errors of the test of hypothesis.

Q.4. (JNTUH 2019, 2 marks): A random sample of size 100 has a standard deviation of 5. What can you say about maximum error with 95% confidence?

Q.5. (JNTUH 2019, 2 marks): Define central limit theorem

Q.6. (JNTUH 2019, 2 marks): Define Type I and Type II errors.

Q.7. (JNTUH 2019, 2 marks): Explain one way classification of ANOVA.

Q.8. (JNTUH 2018, 5 marks): Discuss critical region and level of significance with example

Q.9. (JNTUH 2018, 5 marks): Explain why the larger variance is placed in the numerator of the statistic F. Discuss the application of F-test in testing if two variances are homogenous.

Q.10. (JNTUH 2015, 2018, 5 marks): A sample of 11 rats from a central population had an average blood viscosity of 3.92 with a standard deviation of 0.61. Estimate the 95% confidence limits for the mean blood viscosity of the population.

Answer: df = 11 - 1 = 10; $t_{0.5}$ at df 10 = 2.23

S = SD = 0.61; n = 11, $\bar{x}$ = 3.92

Confidence limits are $\bar{x} \pm t_{0.5}\left(\frac{S}{\sqrt{n}}\right) = 3.92 \pm 2.23\left(\frac{0.61}{\sqrt{11}}\right) = (3.51, 4.33)$

Q.11. (AKTU 2018, 2020, 7 marks): Find the measure of Skewness and kurtosis based on moments for the following distribution and draw your conclusion

Marks	5-15	15-25	25-35	35-45	45-55
No. of students	1	3	5	7	4

Q.12. (AKTU 2019, 3.5 marks): A die is thrown 276 times and the results of those are given below:

No. appeared on the die	1	2	3	4	5	6
Frequency	40	32	29	59	57	59

Test whether the die is biased or not. [Tabulated value of χ^2 at 5% level of significance for 5 degree of freedom is 11.09].

Answer: Solved in this module.

Q.13. (GTU NM 2020, 7 marks): Five coins are tossed 3200 times and the following results are obtained:

No. of heads	0	1	2	3	4	5
Frequency	80	570	1100	900	500	50

If χ^2 for 5 d.f at 5% level of significance be 11.07, test the hypothesis that the coins are unbiased.

Answer: Solved in this module.

Q.14. (GTU NM 2020, 5 marks): In an experiment on immunization of cattle from tuberculosis the following result were obtained:

	Affected	Unaffected
Inoculated	12	26
Not inoculated	16	6

Examine the effect of vaccine in controlling the susceptibility to tuberculosis. (Given that for 5% level of significance χ^2 for 1 d. f is 3.841, for 2 d. f. is 5.99 and for 3 d. f is 7.815).

Answer: Solved in this module.

Q.15. (GTU NM 2020, 5 marks): The heights of 9 males of a given locality are found to be 45, 47, 50, 52, 48, 47, 49, 53, 51 inches. Is it reasonable to believe that the average height is differ significantly from assumed mean 47.5 inches? (Given that for 5% level of significance t for 7 d. f is 2.365, for 8 d. f. is 2.306 and for 9 d. f. is 2.262)

Answer: Solved in this module.

Q.16. (BPUT 2020, 6 marks): In a sample of 1000 people in Odisha, 540 are rice eater and rest are wheat eaters. Can we assume that both rice and wheat are equally popular in this state at 1% level of significance?

Answer: A similar problem is solved in this module.

Q.17. (BPUT 2020, 16 marks): In a sample of 8 observations, the sum of squared deviation of items from the mean was 94.5. In other samples of 10 observations, the value was found by 101.7. Test whether the difference is significant at 5% level of significance?

Answer: Solved in this module.

Q.18. (JNTUH 2018, 10 marks): A sample of 400 items is taken from a normal population whose mean is 4 and variance 4. If the sample mean is 4.45, can the samples be regarded as a simple sample?

Answer: Solved in this module.

Q.19. (JNTUH 2018, 5 marks): In a sample of 600 students of a certain college 400 are found to use ball pens. In another college from a sample of 900 students 450 were found to use ball pens. Test whether two colleges are significantly different with respect to the habit of using ball pens?

Answer: $z = 6.38$ Reject H_0.

Hint: Testing of Significance for Difference of Proportions. See module for formula.

Q.20. (JNTUH 2019, 5 marks): Assuming that $\sigma = 20.0$, how large a random sample be taken to assert with probability 0.95 that the sample mean will not differ from the true mean by more than 3.0 points?

Answer: Solved in this module.

Q.21. (JNTUH 2019, 5 marks): A normal population has a mean of 0.1 and standard deviation of 2.1. Find the probability that mean of a sample of size 900 will be negative.

Answer: Solved in this module.

Q.22. (JNTUH 2009, 2010, 2019, 5 marks): Find 95% confidence limits for the mean of a normality distributed population from which the following sample was taken 15, 17, 10, 18, 16, 9, 7, 11, 13, 14.

Answer:

$$\overline{x} = \frac{15+.....+14}{10} = 13$$

$$s^2 = \frac{\sum (x_i - \overline{x})^2}{n-1} = \frac{1}{9}\left([15-13)^2 + (17-13)^2 + ... + (14-13)^2\right) = 40/3$$

$z_{0.5}$ = 1.96

Confidence limits are $\overline{x} \pm z_{0.5}\left(\frac{S}{\sqrt{n}}\right) = 13 \pm 1.96\left(\frac{3.65}{\sqrt{10}}\right) = (13 \pm 2.26) = (10.74, 15.26)$

Q.23. (JNTUH 2019, 5 marks): In a random sample of 60 workers, the average time taken by them to get to work is 33.8 minutes with a standard deviation of 6.1 minutes. Can we reject the null hypothesis μ =32.6 minutes in favour of alternative null hypothesis $\mu > 32.6$ at $\alpha = 0.025$ level of significance.

Answer: Solved in this module.

Q.24. (JNTUH 2019, 5 marks): The mean life of a sample of 10 electric bulbs was found to be 1456 hours with S.D. of 423 hours. A second sample of 17 bulbs chosen from a different batch showed a mean life of 1280 hours with S.D. of 398 hours. Is there a significant difference between the means of two batches?

Answer: Solved in this module.

Q.25. (UTU 2013, 2016, 10 marks): The following frequency distribution gives the frequencies of seeds in a pea breeding experiment:

Round & yellow	Wrinkled & yellow	Round & green	Wrinkled & green	Total
315	101	108	32	556

Theory predicts that the frequencies should be 9:3:3:1. Examine the correspondence between theory and experiment.

Answer: Solved in this module.

Q.26. (UTU 2013, 5 marks): Before an increase in excise duty on tea,-800 people out of 1000 persons were found to be tea drinkers. After an increase in duty 800 persons were known to be tea drinkers in a sample of 1200. Do you think that there has been significant decrease in the consumption of tea after the increase in excise duty?

Answer: Solved in this module.

Q.27. (UTU 2015, 5 marks): The weight of a drug produced by Ganga Pharmaceutical Co. follows normal distribution. The specified variance of the weight of the drug of this population is 0.25 mg. The quality engineer of the firm claims that the variance of the weight of the drug does not differ significantly from the specified variance of the weight of the drug of the population. So, the purchase officer of the Alpha Hospital who places order for that drug with the Ganga Pharmaceutical Co. has selected a random sample of 12 drugs. The variance of the weight sample is found to be 0.49 mg. Verify the intuitions of the quality manager of Ganga Pharmaceutical Co. at a significance level of 0.10 using Chi Square. (Tabulated chi square value is 19.675 with 11 dof).

Answer: Solved in this module.